The Human Cause-of-Death Database (HCD): Aims, Data and Methods

Inna Danilova, Markéta Pechholdová, Dmitri Jdanov, France Meslé, Jacques Vallin, Domantas Jasilionis, Vladimir M. Shkolnikov

About the Human Cause-of-Death Database Project

The Human Cause-of-Death Database (HCD) is a joint project of the French Institute for Demographic Studies (INED) in Paris, France and the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, based at the MPIDR. The HCD provides free and user-friendly access to coherent time series of cause-specific mortality for researchers, students, journalists, policy analysts, and others interested in analysis of cause-of-death patterns. In contrast to other existing databases on causes of deaths, we provide time series with causes of death classified according to a constant (fixed) list/classification of causes of death.

Our goal is to include the reconstructed data series based on the 10th revision of ICD. Not all countries use ICD-10 at the most detailed (4 digit) level. For comparability purposes, we provide mortality data classified according to intermediate and short lists of causes of death which are standardized for all countries. In addition, data classified according to a more or less "detailed list" (most detailed list of causes of death available for the country), is also provided, according to available data in each country.

The Human Cause-of-Death Database provides access to the following types of data: death counts, age-specific death rates, crude death rates (CDR), standardized death rates (SDR), population exposures, birth counts, and documentation. Death counts and death rates are provided by three cause-of-death lists (full, intermediate, and short) with redistributed ill-defined causes of death and death at unknown age.

The following features make the HCD particularly attractive to its users:

- Continuous data series with constant cause-of-death classification;
- Availability of basic summary (aggregated by age) indicators;
- Detailed documentation;
- Free and easy access to all data;
- The uniform and easy to use format of HCD data files.

We are following as much as possible four guiding principles: comparability, flexibility, accessibility, and reproducibility. We hope that HCD will facilitate comparative analysis of cause-specific mortality data across countries and regions.

The first version of the Database was launched on 29 March 2016 at http://www.causesofdeath.org. Currently the database features cause-specific mortality data for 16 countries (Table 1).

Table 1. Populations currently included in the HCD, with the range of years covered by cause-specific mortality data.

Country	Period
Belarus	1965 - 2010
Czech Republic	1994 - 2013
England & Wales	2001 - 2013

Estonia	1955 - 2012
France	2000 - 2011
Germany	1998 - 2013
Japan	1995 - 2013
Latvia	1956 - 2012
Lithuania	1956 - 2012
Moldova	1965 - 2012
Poland	1970 - 2013
Romania	1980 - 2012
Russia	1965 - 2014
Spain	1980 - 2012
Ukraine	1965 - 2013
USA	1999 - 2013

Motivation for staring the Human Cause-of-Death Database Project

Cause-of-death time series are severely disrupted by periodical changes in the disease classifications. This therefore limits mortality analysis making it possible to analyze cause-specific time-trends only for a short period (covered by the same classification) or only for broad groups of causes of death. Existing international databases do not fix this problem. The information provided by the WHO database is fully ruled by successive ICD revisions and reflect their discontinuity. Eurostat data series are more consistent over time, but for quite a short period of time (from 1994 at best) and for too large groups of causes.

Figure 1. Leukemia mortality trend in Russia before and after reconstruction (age-standardized rates per 100 000)

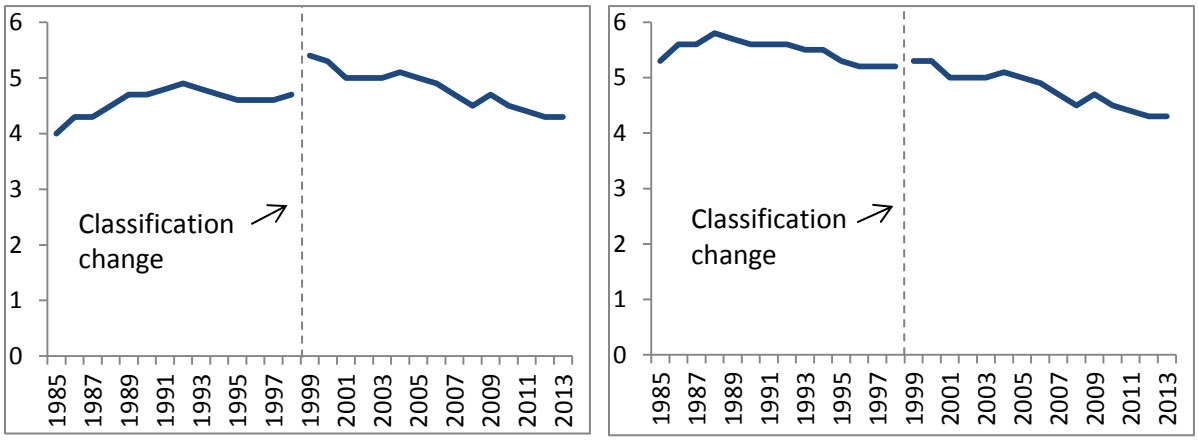


Figure 1 presents age-standardized mortality rates from leukemia in Russia. Introduction of ICD-10 in 1999 caused a noticeable rupture in series (left panel). As a result, death rates from leukemia before and after 1999 cannot be compared directly. In the HCD we overcome the problem of discontinuities in cause-specific mortality data providing series reconstructed in accordance with the most recent version of the classification in use. The right panel of Figure 1 presents the dynamics of age-standardized mortality rates from leukemia in Russia as they are provided by the HCD, with the rates before 1999 reconstructed in accordance with the new classification.

The reconstruction procedure

To reconstruct consistent series, it is necessary to establish transition coefficients between items of two successive classifications, in order to redistribute deaths classified according to the old classification into items of the new classification. When bridge coding (double classification of a sample of deaths simultaneously into the old and new classification) has been performed at a detailed level, transition coefficients can be inferred directly from the results, but there are only two countries in the HCD database where this has been done (and only for the last transition, i.e. from ICD-9 to ICD-10), namely England and Wales and the U.S.A. For the other transitions in the two countries and for all transitions in other countries, coherent time series can be reconstructed by producing ex-post double coding. To perform such kind of reconstruction we use as a guideline in the HCD the method developed at INED in the 1980s. But the work was tailored to each country independently.

For each classification change, the method of reconstruction comprises three steps (Vallin and Meslé, 1988, 1998; Meslé and Vallin, 1996):

- Setting up one correspondence table which lists, for each item of one classification, all the items of the other one that are a priori equivalent in terms of medical content.
- Building fundamental associations of items that identify the smallest possible number of items containing the same medical contents in both classifications and testing the consistency of the associations over time using a statistical test (Barbieri, Chung, and Boe, 2008; Camarda, Pecholdová, and Meslé, 2015).
- Setting up ex-post double-coding according to the structure of fundamental associations, to finally obtain transition coefficients.

The results derived from the medical logic of the classification rules have to be checked statistically, to detect and solve any remaining breaks in the series. Such checks are carried out by age group and sex.

In addition, national statistical offices make occasional changes independent of the official revisions of the classification. To address this problem, the statistical continuity of the series over time is systematically verified and any artificial disruption dealt with appropriately.

Finally country- and time-specific methods are used to deal with ill-defined causes (Ledermann, 1955; Vallin and Meslé, 1988).

Overview of the Human Cause-of-Death Database

Our goal is to include the reconstructed data series based on the 10th revision of ICD. Not all countries use ICD-10 at the most detailed (4 digit) level. For comparability purposes, we provide mortality data classified according to intermediate and short lists of causes of death which are the same for all countries. In addition, data classified according to a more or less "detailed list", is also provided, according to available data in each country.

At present, we have several countries with short data series started at the moment when ICD-10 was introduced in the country and restricted only to short and intermediate lists of causes. We are working on data reconstruction for these countries, and in near future these data will be replaced by longer and reconstructed time series with constant classification of causes.

We provide complete documentation of the data available through this site. Documentation that is specific to an individual population (including data sources) is provided through links within

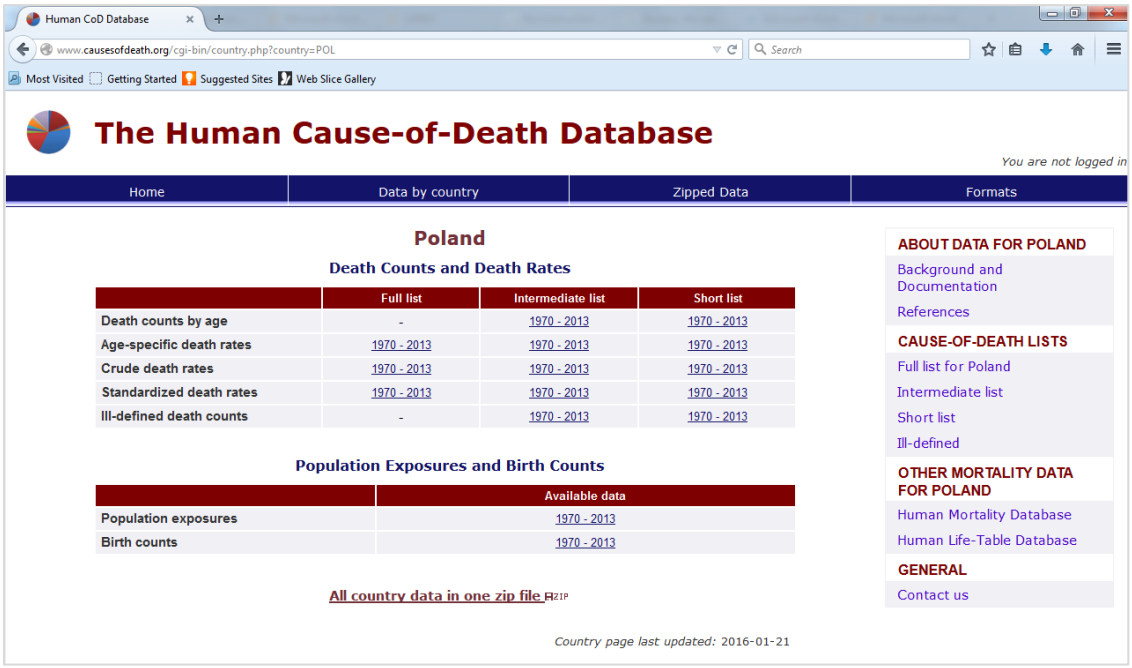
each country section, documentation which is relevant to the whole project is available via links on the main page.

In addition to re-classification of causes of death classified according to old ICD revisions, we redistribute ill-defined causes of death and adjust infant mortality (if needed). Deaths of unknown age are redistributed proportionally into known ages for each cause of death. A detailed description of data adjustments for every country is provided in the country-specific documentation.

For every country in the HCD, original death counts are a result of individual research work and provided as output for the database by individual researchers (country specialists).

For calculation of age-specific death rates, crude death rates, and (age) standardized death rates population exposures are taken from the Human Mortality Database. For countries that are not included in the HMD, we use the standard HMD methodology to produce our own population exposures.

Figure 2: Example of a country data page



Future plans

Our goal is to establish the HCD as a respected and widely used source of high-quality data on mortality by causes of death. We hope the Database will become one of the key resources for cross-country comparative studies. The database will be regularly updated.

For a few countries in the project the reconstruction work is still in progress. Once the reconstructed series will be available they will enter the Human Cause-of-Death Database progressively.

We also aim to improve the website and data presentation, following the feedback from the HCD users.

References

1. Barbieri, M., Chung, R., & Boe, C. (2008). *Automating the redistribution of deaths by cause over ICD changes.* Second Human Mortality Database Symposium, Max Planck Institute for Demographic Research, Rostock, Germany, 13-14 June 2008.
2. Camarda, C.G., Pechholdová, M. & Meslé, F. (2015). *Cause-specific senescence: classifying causes of death according to the rate of aging.* 80th Annual Meeting of the Population Association of America. San Diego (USA), May 2015. http://paa2015.princeton.edu/uploads/153074
3. Ledermann, S. (1955). *La répartition des décès de cause indéterminée.* Revue de l'Institut international de statistique, 23 (1–3), 47–55.
4. Meslé, F., & Vallin, J. (1996). *Reconstructing long-term series of causes of death*. Historical Methods, 29 (2), 72–87.
5. Vallin, J., & Meslé, F. (1988). *Les causes de décès en France de 1925 à 1978* (Travaux et Documents, No.115, 608 p.). Paris: INED/PUF.
6. Vallin, J., & Meslé, F. (1998). *Comment suivre l'évolution de la mortalité par cause malgré les discontinuités de la statistique. Le cas de la France de 1925 à 1993.* In G. Pavillon (Eds.), Enjeux des classifi cations internationales en santé (Questions en santé publique, pp. 113–156, 220 p.). Paris: Éditions INSERM.